

Data Journalism

Michael Friendly
 Psych 6135

<http://euclid.psych.yorku.ca/www/psy6135>

Topics

- What is DDJ?
- Some early examples
- Modern examples
 - election reporting
 - health statistics
 - social media
- Best examples??
- Now trending

What is data journalism?

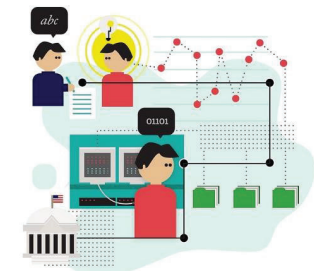
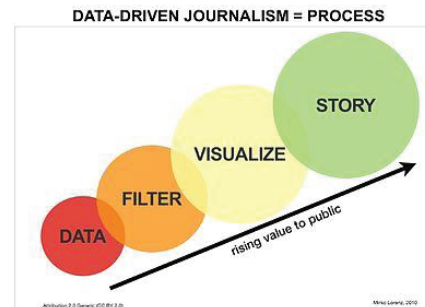
- **Data journalism** reflects the increased role of numerical data for reporting in the digital era. Data driven or DDJ more useful
 - Reflects increased interaction between journalists and {visualization design, computer science and statistics}.
- Aspects:
 - **Database analysis:** FOI requests on topics like crime, racial discrimination, judicial sentencing, toxic pollution, ...
 - **Visualization:** infographic ← → data vis., emphasis on the **story**
 - Increasing use & development of data vis. **software tools**
 - analysis: R, python, ...
 - visualization libraries: R, D3, javascript, ...
 - Increasing use of **dynamic updating** of data sources
 - election polls / prediction
 - disease tracking

What is data journalism?

To an outsider, it is often hard to tell concretely what #DDJ is, or is not

Two motivational infographics of #DDJ as a process

- ❖ How does it actually work?
- ❖ What makes it successful?



Bill Dedman: *The Color of Money*

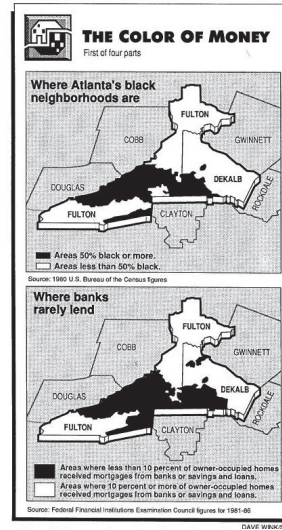
Bill Dedman of *The Atlanta Journal-Constitution* analyzed racial discrimination by banks and other lenders in middle-income black neighborhoods of Atlanta, GA.

In a series of articles, May 1-4, 1988, *The Color of Money*, he showed that although banks made loans in even the poorest white neighborhoods, they did not lend in middle-class or more affluent black neighborhoods.

This work won the Pulitzer prize in investigative reporting in 1989.

It helped establish the idea of computer-assisted reporting and the importance of compelling visualizations in telling the story.

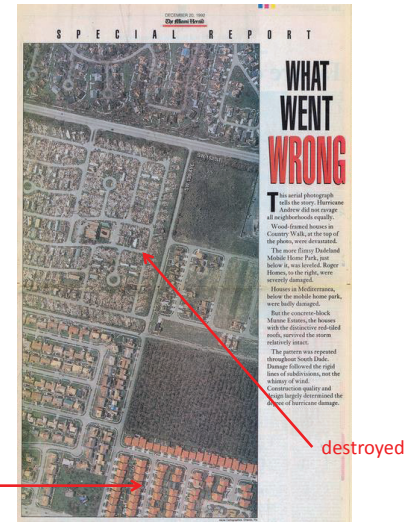
See: <http://powerreporting.com/color/> for all the articles and images



5

Hurricane Andrew

- In Aug. 1992, a category 5 hurricane hit Miami-Dade county
- Huge numbers of houses were destroyed, but this varied by neighborhood
- Investigative reporters at the *Miami Herald* discovered why:
 - Building codes and inspection became lax after 1980.
 - Older homes survived.
- This report by Liz Balmaseda et al. won the Pulitzer Prize in 1993.



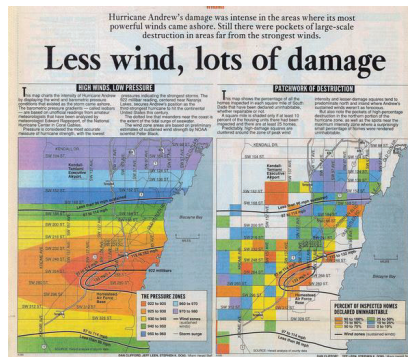
undamaged

destroyed

Source: <https://www.flickr.com/photos/juggernautco/sets/72157607210036175/>

Hurricane Andrew

- The report in the *Miami Herald* contained many detailed analysis maps
- There was compelling evidence for the cause of destruction.
- The combination of investigative reporting, analysis and graphics was seen as powerful



7

Predicting elections

- In 1962, John Tukey and a statistical team joined with NBC to attempt to predict the congressional election
- Over time, they developed new methods to combine **past results** with **exit polls** and **current returns** to give up-to-date estimates and measures of uncertainty as the numbers came in.
- Key idea: "borrowing strength"
- Now known as "hierarchical Bayes"



John Tukey (front, left) and Richard Link (front, right) as part of the NBC Election Day team in Rockefeller Plaza Studios with RICANBC staff in the background, 1962. Photo courtesy of the American Philosophical Society. Tukey Archives

NBC execs were freaked out when the Tukey team began to call elections quite early with 95%, then 97.5% certainty. They waited until it reached 99%, and missed the scoop.

8

CNN Magic Wall

Direct, interactive visualization of election news on TV took a big jump with the development of the Magic Wall, a large touch-screen display, tied to databases of election results, and with on-screen controls to zoom in/out, compare, highlight, etc.



Video: <https://www.youtube.com/watch?v=rpLHnK-MSy4>

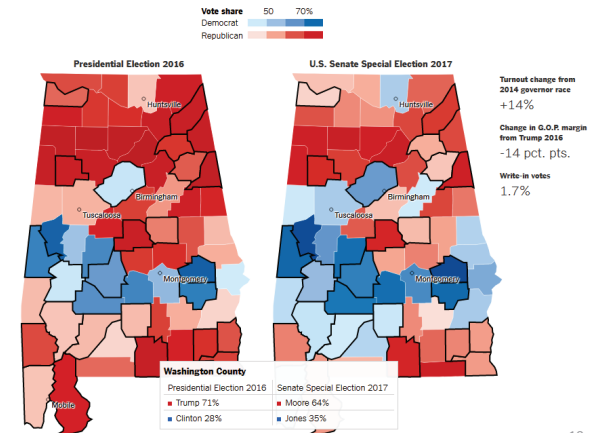
Election reporting: NYT

In the special election for the Alabama U.S. Senate, Dec. 12, 2017, *The New York Times* used dynamic, interactive maps to compare the vote to the 2016 Presidential election

Hovering the mouse over a county showed the details

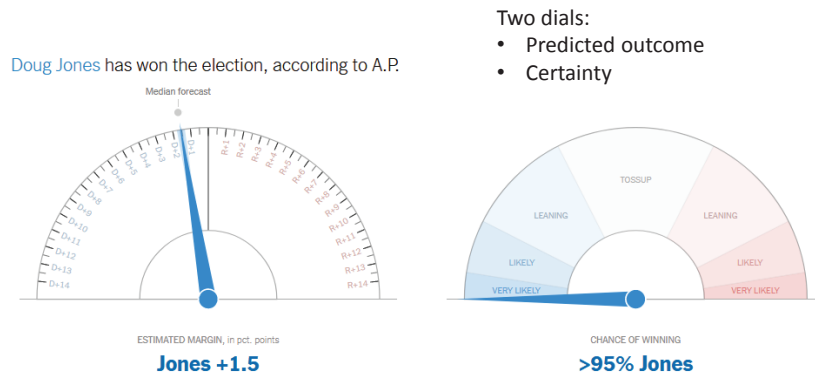
A map-based display is natural & easy

Q: What other graphic might show the comparison better?



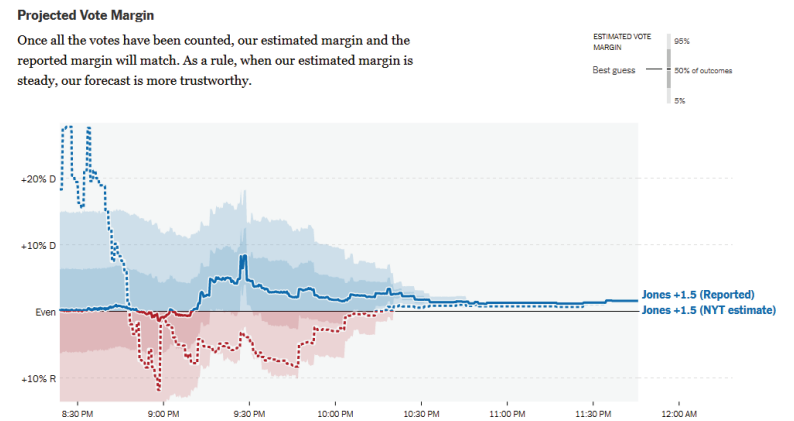
Election reporting: Dashboards

Throughout the evening of Dec. 12, the NYT dynamically updated dashboard displays of the election returns and model-based predictions of the winner



Election reporting: Dashboards

Another dynamically updated chart showed the projected vote margin, together with two levels (50%, 95%) of **uncertainty intervals** for the current estimate.



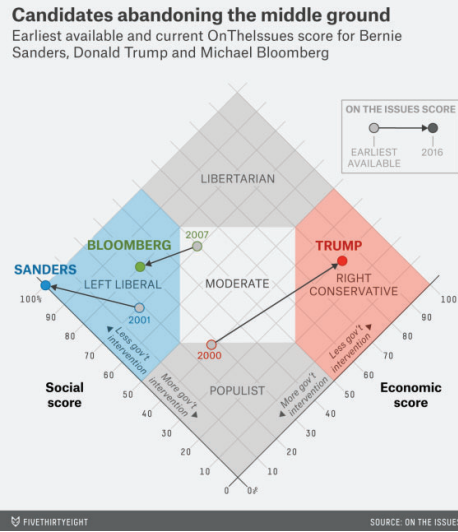
Novel graphic designs

This graph, from fivethirtyeight.com was designed to show how some presidential candidates had shifted positions before the 2016 election.

The axes are a score on **social** and **economic** policy, but they rotated the axes by 45° to create zones related to political thought.

This info graphic is eye-catching and self-explanatory:

- colored/labeled zones
- interpretive labels on axes
- arrows showing movement to extremes



Measles and vaccines

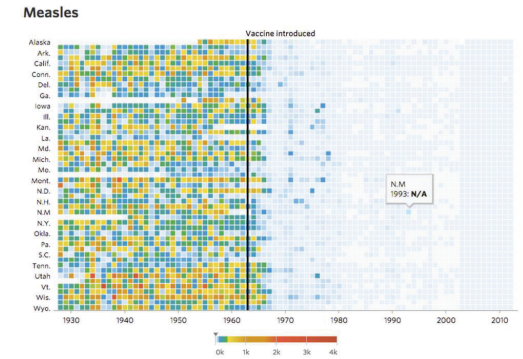
Visualizing the impact of health policy interventions

In 2015 Tynan DeBold & Dov Friedman in the *Wall Street Journal* tried to show the effect of the introduction of vaccination programs in the US states on disease incidence, using color-coded heat maps for a variety of diseases

The long time series ~70 years made this work.

The heat map color scale is not exemplary, but the message is still clear: disease incidence declined after vaccines were introduced.

The images are “interactive,” in the weak sense that tool-tips are shown on mouse movement.

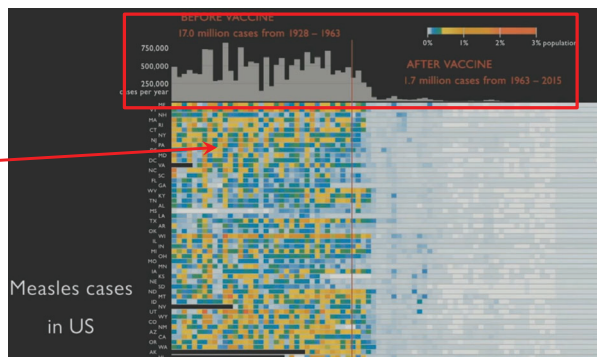


Source: <http://graphics.wsj.com/infectious-diseases-and-vaccines/>

Making the message more explicit

What you should try to show here in an Info graphic is the **overall impact** of vaccination on measles
Ed Tufte did this by adding a histogram at the top showing **total # of cases** by year

With the overall impact made clear, details about individual states can now be explored



Effect ordering & color scales

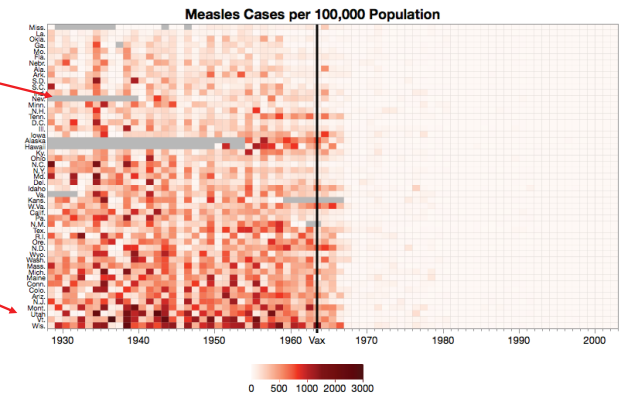
Ordering the states by an overall measure of measles incidence makes it easier to study details.

A uni-dimensional scale of color is more effective than the ad-hoc scale used in the original

This chart, by Xan Gregg also shows missing data explicitly.

Color scales for data graphs is a growing topic.

What's wrong in Wis. & VT?

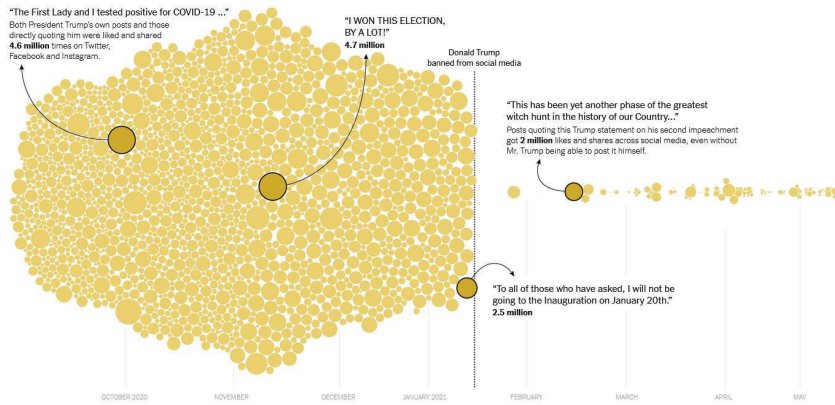


From: <https://community.jmp.com/t5/JMP-Blog/Graph-makeover-Measles-heat-map/ba-p/30550>

Before – After Circle Chart

What Happened When Trump Was Banned on Social Media?

- ❖ Circles are statements by Trump.
- ❖ Circle size ~ likes & shares on social media



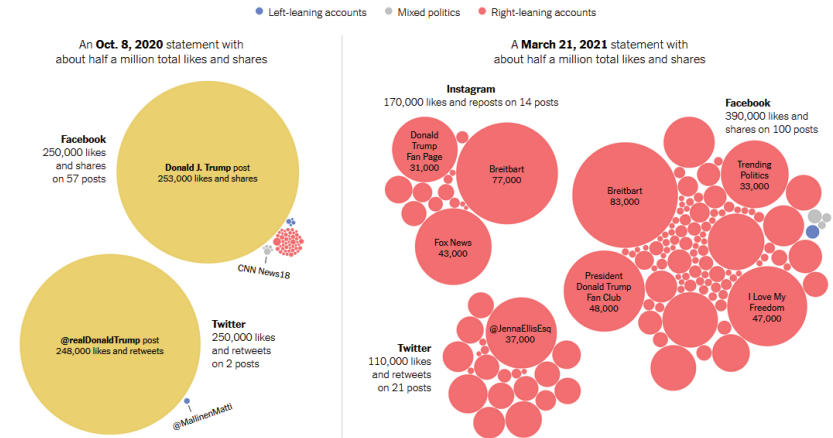
<https://www.nytimes.com/interactive/2021/06/07/technology/trump-social-media-ban.html>

17

Before – After: Who pushed the lies?

How/Who Spread Trump statements before and after the ban

Posts referencing each Trump statement, sized by total number of likes and shares



18

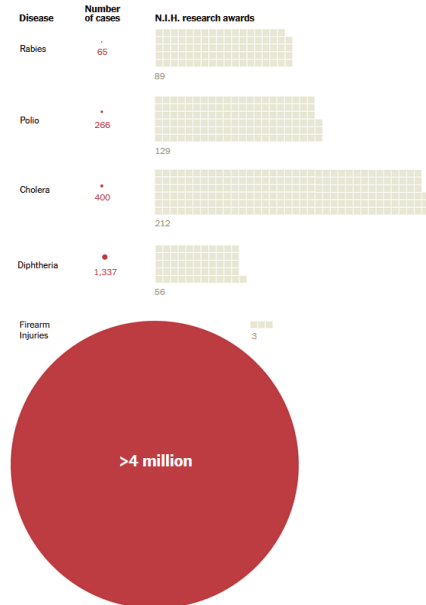
Health & research funding

This powerful NYT chart compares cases of diseases and firearm injuries to the number of N.I.H. research awards.

How well do you like this design?

Is it a fair depiction of the data?

Can you suggest anything different?



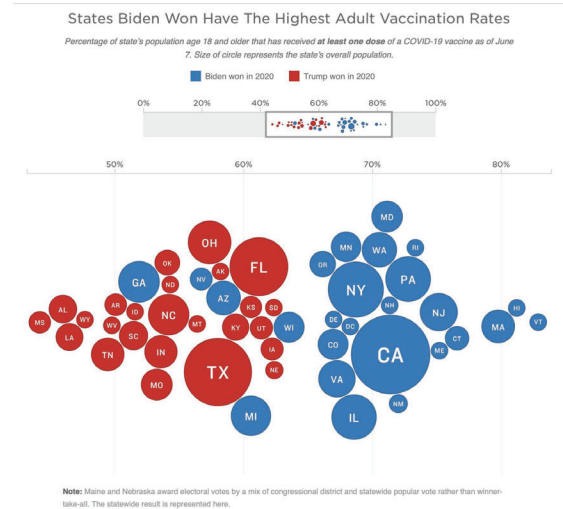
Source: University of Chicago Crime Lab

19

Votes for vs. COVID vaccinations

% of state pop. with at least one dose, vs. who won in 2020

circle size ~ state pop.



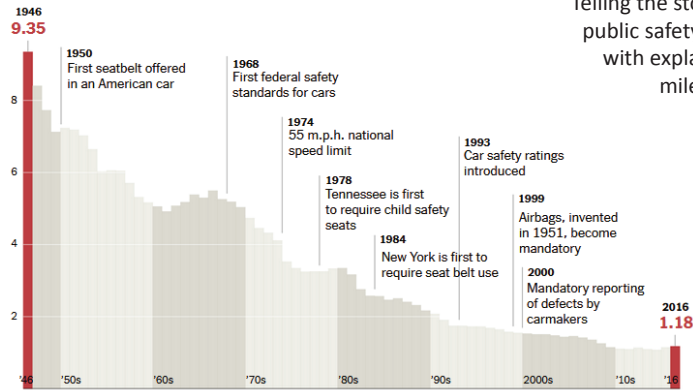
From: <https://twitter.com/kihealy/status/1402691186826039301>

20

Seatbelts

New York Times: An annotated chart of vehicle deaths in relation to seatbelts

Deaths per 100 million vehicle miles traveled



Source: National Highway Traffic Safety Administration

Telling the story of a public safety policy with explanatory milestones

Best examples of DDJ? Ask Twitter

Michael Friendly @datavisFriendly · 22h
1/ Thinking about an upcoming class in my #psyc6135 #datavis course on the topic #datajournalism, with a guest star @Jleerankin . It got me thinking of the idea of "Great moments in Data Journalism": investigative reporting, incisive analyses, compelling graphic displays.

Michael Friendly @datavisFriendly · 22h
2/1 would love ❤️ for nominations on this topic. @albertocairo and others have done some of this, but perhaps this is a topic on which others could contribute their nominations for "Great moments..."

Alberto Cairo @albertocairo · 22h
Twitter #ddj and news #dataviz. contribute with nominations.
I'd do @FiveThirtyEight's: projects.fivethirtyeight.com/redistricting-...
Anything by @ProPublica, like @lenagroeger's appropriately creepy: projects.propublica.org/graphics/worke...
TPTimes's [tampabay.com/projects/2015/...](https://tampabay.com/projects/2015/)

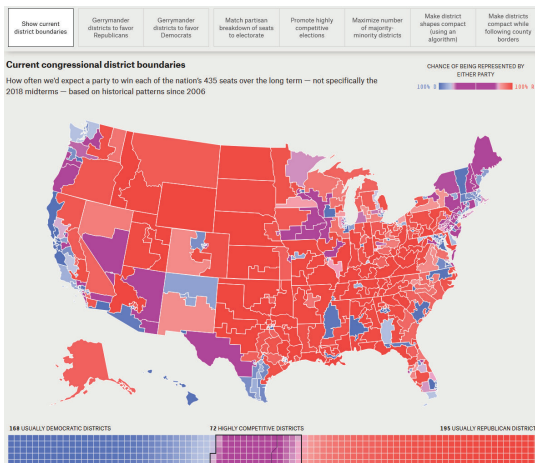
538: Atlas of Redistricting

There's a lot of complaining about gerrymandering, but what should US Congressional districts look like? What would they look like to serve various goals?

This map shows the current boundaries, with expected wins by **Rep** and **Dem**.

Why do I find this app appealing?

From: <https://projects.fivethirtyeight.com/redistricting-maps/>



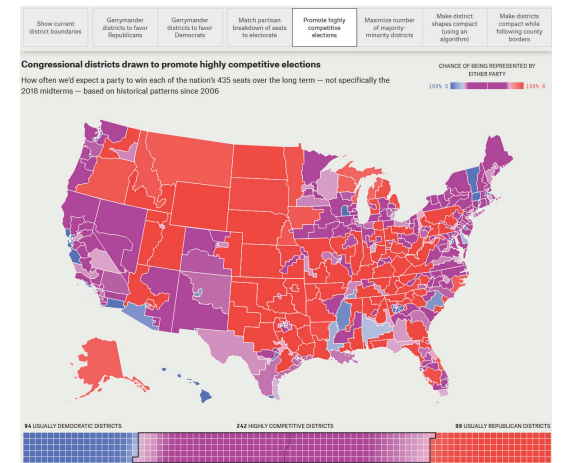
538: Atlas of Redistricting

How could district boundaries be re-drawn to promote more competitive elections?

Election district boundaries are consequential & highly political. They can be changed by state legislatures.

Different scenarios show the changes in representation, w/o a single voter switching parties!

From: <https://projects.fivethirtyeight.com/redistricting-maps/>

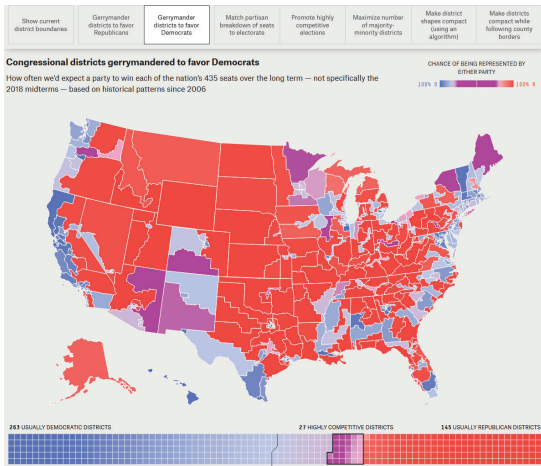


538: Atlas of Redistricting

What if the Democrats could control election districts to favor themselves?

Note that the overall impression of the map is still **Rep.** overall.

Changes in representation take place largely on the east/west coasts



From: <https://projects.fivethirtyeight.com/redistricting-maps/>

WP: Segregation in U.S. Cities

Aaron Williams & Armand Emamdjomeh at the Washington Post studied changes in racial composition and diversity in U.S. cities over 30+ years

They relate these to demographic, policy, and urban/suburban issues using well-designed thematic maps.



From: <https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/>

Our World in Data

An impressive range of social & environmental issues presented by Our World in Data

Data Explorer

We just published our new data explorer on Crop Yields

Explore our new data explorer covering yields and yield gaps for a range of crops across the world.

By Hannah Ritchie — June 08, 2021

Why do we need to know about progress if we are concerned about the world's large problems?

Why have we made it our mission to publish the "research and data to make progress against the world's largest problems"?

By Max Roser — June 07, 2021

The price of batteries has declined by 97% in the last three decades

To transition towards low-carbon energy systems we need low-cost energy storage. Battery costs have been falling quickly.

By Hannah Ritchie — June 04, 2021

The argument for a carbon price

We are paying a price for fossil fuels, but that price is not paid by those that burn the fossil fuels — we need to change that.

By Max Roser — June 01, 2021

Living Planet Index: what does an average decline of 68% really mean?

The Living Planet Index is the biodiversity metric that always claims the headlines. It's often misinterpreted. How should we understand it?

By Hannah Ritchie — May 17, 2021

How the Living Planet project helps us understand how the world's wildlife is changing

Beneath the popular index, the Living Planet database helps us understand where and what animals are deeply threatened, and what animals are thriving.

By Hannah Ritchie — May 17, 2021

Financial Times COVID tracker

John Burn-Murdoch @jburnmurdoch

This series of charts from the Financial Times was hailed for best-practices in the early stages of the pandemic.

- ❖ Log scale make it easy to compare rates of increase
- ❖ Visual meaning for "leveling the curve"
- ❖ Labels for interventions (lockdown)
- ❖ Sensible (relative) origin – days since 10th death

The @FinancialTimes coronavirus death & case trajectory trackers are now FREE TO READ outside the paywall: ft.com/coronavirus-la...

In this morning's update, the US has gone above 470 deaths, bringing it just behind where Iran was at the same stage.



Toronto Star: Trump Lies

Daniel Dale & Ed Tubb at the *Toronto Star* maintained a database of every false claim made by Donald Trump since the Inauguration on Jan. 20, 2017



Donald Trump has said **4557** false things as U.S. president

(as of 3-13-2019)

Interactive filter at: <http://projects.thestar.com/donald-trump-fact-check/>

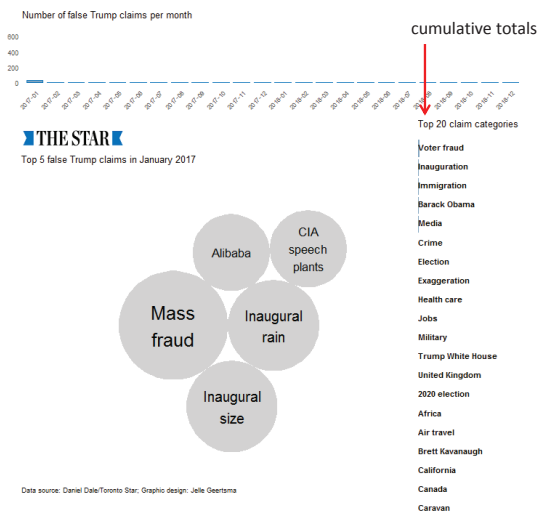
Toronto Star: Trump Lies

Every false claim is documented: place, claim, truth, "slug", categories

number	date	place	claim	truth	slug	repeat	source	category1	category2
1	2017-01-20	Post-inauguration...	"I looked at the rain, which...	The rain began right at the beg...	Inaugu...	Yes	https://www.thestar.com/ne...	Inauguration	Weather
2	2017-01-20	Post-inauguration...	"Even the media said the c...	The major media reported that ...	Inaugu...	Yes	https://www.thestar.com/ne...	Inauguration	Exaggerat
3	2017-01-21	Speech at Centr...	"Honestly, it looked like a ...	The crowd, which may not have...	Inaugu...	Yes	https://www.youtube.com/w...	Inauguration	Exaggerat
4	2017-01-21	Speech at Centr...	"It was almost raining, the ...	Neither of these claims is true...	Inaugu...	Yes	https://www.youtube.com/w...	Inauguration	NA
5	2017-01-21	Speech at Centr...	"So a reporter for TIME ma...	Trump's numbers are well off. H...	TIME co...	Yes	https://www.youtube.com/w...	Media	Exaggerat
6	2017-01-23	Private meeting ...	Trump told Congressional ...	This claim, also reported by nu...	Mass fr...	Yes	http://www.foxnews.com/pol...	Voter fraud	Election
7	2017-01-25	Interview with A...	"In terms of a total audien...	"They" can mean anyone, but n...	Inaugu...	Yes	http://abcnews.go.com/Polit...	Inauguration	Exaggerat
8	2017-01-25	Interview with A...	"No, no, you have to unde...	Trump's victory was not close L...	Election...	Yes	http://abcnews.go.com/Polit...	Election	Exaggerat
9	2017-01-25	Interview with A...	"I think you're demeaning ...	Saying "you" here, Trump wron...	Media	Yes	http://abcnews.go.com/Polit...	Media	Exaggerat
10	2017-01-25	Interview with A...	"With regard to his speech ...	Most of the audience was inde...	CIA spe...	no	http://abcnews.go.com/Polit...	Trump White House	US intell...
11	2017-01-25	Interview with A...	Regarding Healthcare.gov...	Healthcare.gov did not cost \$5 ...	Healthc...	no	http://abcn http://abcnews.go.com/...	are	Exaggerat
12	2017-01-25	Interview with A...	Muir: "I called the author ...	Groveling means "to draw bac...	Grovelin...	no	http://abco Politics/transcript-ab...	jud	NA
13	2017-01-25	Interview with A...	"Now you're telling me Pe...	Trump was trying to use a 2012 ...	Pew fra...	no	http://abco News/press-preside...	jud	Election
14	2017-01-25	Interview with A...	Regarding his false claim ...	These large numbers of illegal w...	Mass fr...	Yes	http://abco story?id=45047602	Chicago	Election
15	2017-01-25	Interview with A...	Regarding voting fraud: "...	There is no evidence of a signifi...	Mass fr...	Yes	http://abcnews.go.com/Polit...	Voter fraud	Chicago
16	2017-01-25	Interview with A...	Regarding voting fraud: "...	There is no evidence of a signifi...	Mass fr...	Yes	http://abcnews.go.com/Polit...	Voter fraud	NA
17	2017-01-25	Interview with A...	"Look, Barack Obama -- if ...	This is a gross mischaracterizati...	Obama	no	http://abcnews.go.com/Polit...	Barack Obama	Voter frau
18	2017-01-25	Interview with A...	"So, look, when President ...	There were not only no homici...	Obama	no	http://abcnews.go.com/Polit...	Barack Obama	Crime

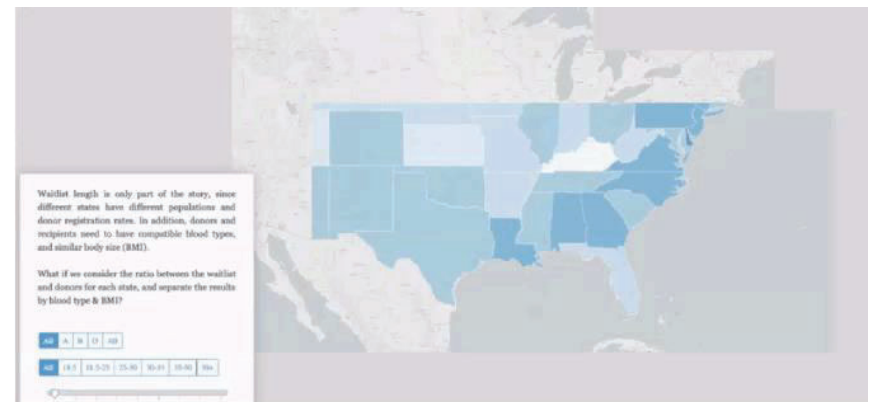
Jelle Geertsma: Trump Lies

Graphic artist Jelle Geertsma designed this animated graphic showing the top 5 claim "slugs" & the top 20 categories over time, by month.



Scrollytelling: The Impatient List

What it's like waiting for a kidney transplant in the U.S.



"The Impatient List", an example of Scrollytelling by Gupta et al.

Now trending ...

- **Digital publishing and online, interactive media**
 - Many news organizations have interactive graphic developers or entire online departments
 - A revenue stream responding to less from print & ads
- **Story telling frames a wider enterprise**
 - A focus for data visualization
 - New online forms: scrollytelling
- **Open data**
 - Maps-GIS (Google, Open Street maps)
 - Gov't data: census, health, crime, ...
 - NGO initiatives: Our World in Data
- **Software infrastructure**
 - Javascript as base for online, interactive methods
 - Libraries, frameworks building on this (D3.js, Processing, ...)